Using Tree Diagrams without Numerical Values in Addition to Relative Numbers Improves Students' Numeracy Skills: A Randomized Study in Medical Education

Hendrik Friederichs, MD, Sandra Ligges, Dipl-Stat, Anne Weissenstein, MD

Background. *Physicians and medical students may lack sufficient numeracy skills to make treatment decisions, interpret test results, and practice evidence-based medicine. We evaluated whether the use of a tree diagram without numerical values as an aid for numerical processing might improve students' test results when dealing with percentages.* ***Methods.*** *A prospective randomized study was carried out with 102 third-year students. Participants received 3 diagnostic test problems and were asked to determine positive predictive values. The information in these tests was expressed either in (1) natural frequencies, (2) conditional probabilities, or (3) conditional probabilities with a tree diagram without numbers.* ***Results.*** *Ninety-eight (96.1%) complete data sets could be obtained. The group working with natural frequencies achieved significantly higher test results (n = 29, mean score: 1.1, P = 0.045) than the group working with conditional probabilities (n = 34, mean score: 0.56). The students who were given a tree diagram in addition to conditional probabilities (n = 35, mean score: 1.26) also achieved significantly better scores than the group with conditional probabilities alone (P = 0.008). The difference between the group who had received natural frequencies and the group working with conditional probabilities and the tree diagram was not significant.* ***Conclusions.*** *We suggest the use of a tree diagram as a visual aid when dealing with diagnostic tests expressed in conditional probabilities.* ***Key words:*** *numeracy; tree diagram; medical education; conditional probabilities.* ***(Med Decis Making 2014;34:253–257)***

Physicians as well as medical students have repeatedly been shown to have poor understanding of statistics and quantitative aspects of research.[1–6] Especially in applying Bayesian inferences that are used to estimate probabilities of a hypothesis in the face of new data, medical students and physicians lack skills.[7–10] A possible way for students to improve their numeracy skills is through curricular courses addressing this issue. However, in a study by Rao and Kanter,[11] the authors discovered that numeracy is seldom a part of medical curriculum. In addition, several studies have indicated that students generally dislike courses involving mathematics, numeracy, or statistics.[2,12] Finally, courses do not always improve students' diagnostic performance.[13]

A different approach addressing this issue has been introduced by Hoffrage and Gigerenzer,[14] who postulated that using natural frequencies instead of conditional probabilities (e.g., sensitivity and specificity) would facilitate diagnostic insight. This is especially important when understanding the probability of joint occurrence of 2 events, in contrast to the probability of single events, for which percentages may be more suitable. Unfortunately, in most medical textbooks and curricula, conditional probabilities still predominate. That is why the members of the medical faculty in Muenster, Germany, have continued to search for a way to facilitate students'

Received 16 April 2013 from University of Muenster, Studienhospital, Muenster, Germany (HF, AW); and University of Muenster, Institute of Biostatistics and Clinical Research, Muenster, Germany (SL). No financial support was provided for this study. Revision accepted for publication 15 August 2013.

Address correspondence to Anne Weissenstein, Studienhospital der Medizinischen Fakultät Münster, Malmedyweg 17-19, 48149 Muenster, Germany; telephone: + 49 251 83-53101; fax: + 49 251 83-53199; e-mail: anne.weissenstein@gmail.com.

DOI: 10.1177/0272989X13504499

Table 1 Three Case Scenarios with Varying Given Variables and One Target Variable

	Given Variables	Target Variable
Mammography	Prevalence, sensitivity, and 1-specificity	Positive prediction
AIDS counseling	Prevalence, sensitivity, and specificity	Positive prediction
Back pain	Prevalence, 1-sensitivity, and specificity	Positive prediction

inferences when using conditional probabilities. The primary goal of our study was to evaluate whether providing a (blank, general) tree diagram in addition to conditional probabilities can enhance students' numeracy skills compared with conditional probabilities alone or natural frequencies.

METHODS

Study Design and Participants

We performed a prospective randomized study comparing students' numeracy skills in 3 case scenarios expressed in different frequency formats. The participants were 102 third-year students (first clinical semester) at the medical school of the University of Muenster. Each participant was assigned to 1 of the 3 test formats by lot. Participants attempted the case scenarios voluntarily and anonymously. Informed consent was provided by all participants before the study. In light of the aforementioned study design, the Ethics Committee of the Chamber of Physicians Westfalen-Lippe and the Medical Faculty of the Westfalian Wilhelms University Muenster waived requirements for an ethical approval procedure.

Case Scenarios

Three case scenarios were constructed, the first 2 according to the scenarios by Gigerenzer and Hoffrage[15] and the third by the authors based on a meta-analysis in the Cochrane database.[16]

In each case, we presented information concerning disease prevalence and either conditional probabilities (sensitivity and specificity or their complements) or natural frequencies (numbers of patients out of a fixed number with positive or negative tests) (Table 1). The fixed number of patients was chosen in a way that it was possible to calculate with natural frequencies for individuals with positive and negative tests. The task was to give a probability estimate for the positive predictive value. An example of a case presented with conditional probabilities and natural frequencies is given.

Conditional probabilities:

> For early detection of breast cancer, women over 50 years can participate in a mammography screening. About 1% of these women have breast cancer (prevalence). On a mammogram, a breast cancer can be detected with 90% probability (sensitivity). The probability of a false-positive result in a healthy woman is 9%. What is the probability a woman has breast cancer if she has a positive finding in the mammography? (Please indicate the likelihood as a percentage; do not use any decimals or commas.) _________%

Natural frequencies:

> For early detection of breast cancer, women over 50 years can participate in a mammography screening. Of these women, about 10 in 1000 have breast cancer (prevalence). On a mammogram, 9 of these 10 women would get a positive result (sensitivity). Of 990 healthy women, 89 women also get a positive result. What is the probability a woman has breast cancer if she has a positive finding in the mammography? (Please indicate the likelihood as a percentage; do not use any decimals or commas.) _________%

Students received the scenarios presented either in natural frequencies, in conditional probabilities, or in conditional probabilities with a tree diagram without numerical values as an aid for numerical processing (Figure 1).

Outcome Measures

Outcome measures were the mean test scores students achieved in the 3 scenarios. For each correctly answered case, 1 point was given; therefore, participants could obtain 0, 1, 2, or 3 points. In addition, baseline characteristics (age, gender) were recorded and students provided a self-assessment of their statistical skills and their statistical education.

- How confident are you in calculating percentages?
- How confident are you with statistics?
- What course did you attend in sixth form? Basic or advanced?

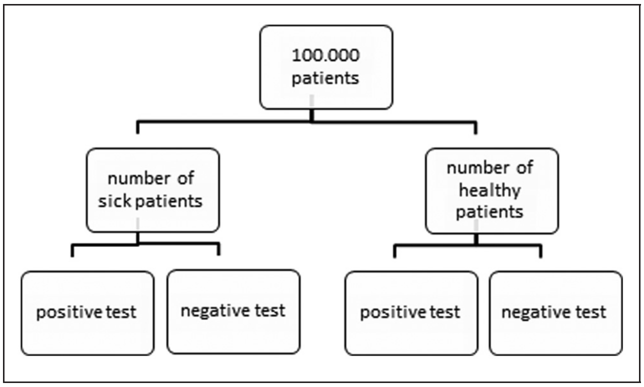

Figure 1 *Tree diagram provided in addition to conditional probabilities.*

- Did the course cover statistics?

The first 2 questions were assessed using a 6-point category rating scale analogous to the German school grading system (1 = very good, 2 = good, 3 = average, 4 = below average, 5 = deficient, 6 = poor).

Data Processing and Analysis

Data were analyzed with R 3.0.1 (R Core Team, 2013) and IBM Statistics SPSS 21. First, using analysis of variance and chi-square tests, we checked whether the baseline characteristics and self-assessment of statistical skills and statistical education were balanced among the groups.

To compare the students' performance in the 3 different presentations, we compared scores of the students among groups using the Kruskal-Wallis test for all groups simultaneously and Mann-Whitney *U* tests for the pairs of groups. All significance tests were carried out at a 2-sided significance level of $\alpha = 0.05$.

Because we felt that only a large effect size in the comparisons would be meaningful, we sought to obtain 28 participants per group in order to have 80% power to detect an effect size of $r = 0.5$.[17]

RESULTS

Participants Characteristics

Ninety-eight complete data sets were obtained from 102 students (return rate 96.1%.) As Table 2 reveals, the baseline characteristics (age and gender) of students as well as self-assessment of statistical skills and statistical education were balanced among the 3 randomized intervention groups.

Effect of Presentation

Figure 2 shows the distribution of scores among students in the 3 groups. The proportion of students achieving the minimum score (0 points) is highest for the conditional probabilities group, at the medium position for the natural frequencies group, and lowest for the group presented with the tree diagram in addition to conditional probabilities. The proportion of students achieving the maximum score (3 points) follows the opposite pattern. The Kruskal-Wallis test shows a significant difference among the scores of the different intervention groups (P = 0.022). The P values of the Mann-Whitney U tests are 0.008 for the pair of tree diagram with conditional probabilities and conditional probabilities alone, 0.045 for the pair of natural frequencies and conditional probabilities alone, and 0.588 for the pair of natural frequencies and tree diagram with conditional probabilities. This effect was also reflected in the group-wise mean scores of 0.56 (conditional probabilities), 1.1 (natural frequencies), and 1.26 (tree diagrams with conditional probabilities).

DISCUSSION

In all 3 groups the obtained mean scores were low, which is not surprising given the previously discussed low numeracy skills of medical students.[7,8] However, the results of our study show that simply by providing the tree diagram without numerical values, test scores could be improved significantly. This is consistent with other studies which found that visual aids in addition to numerical information help people to make more accurate assessments.[18–20]

In past studies, visual aids representing numerical values were used. Graphs, icon arrays, or other visual formats reporting data were presented in addition to the numbers. However, this type of presentation is infeasible for every medical situation, as too many studies would have to be processed. In contrast, our tree diagram without values gives an overview of all negatively and positively tested subjects, healthy as well as ill, in the general population. Garcia-Retamero and Galesic[19] showed that participants profit more from visual displays when the overall population at risk is depicted than when only sick individuals are shown. The tree diagram is a general tool in which the data can be entered and calculated by students.

Courses may not necessarily enhance students' numeracy skills,[13] and when they do the courses

Table 2 Baseline Characteristics of the Groups

	Natural Frequencies (*n* = 29)	Conditional Probabilities (*n* = 34)	Conditional Probabilities Plus Diagram (*n* = 35)	*P* Value
Mean age, years	24.03 ± 3.8	23.5 ± 2.8	23.5 ± 2.9	0.992
Female gender, % (*n*)	51.7 (15)	47.1 (16)	57.1 (20)	0.740
Confidence in calculating percentages (6 = lowest to 1 = highest)	2.41	2.56	2.66	0.618
Confidence in statistics (6 = lowest to 1 = highest)	3.59	3.65	3.71	0.844
Statistical education in sixth form, basic or advanced course (advanced course %)	35.7	42.4	45.5	0.754
With or without statistics (with statistics %)	41.7	44.8	46.4	0.960

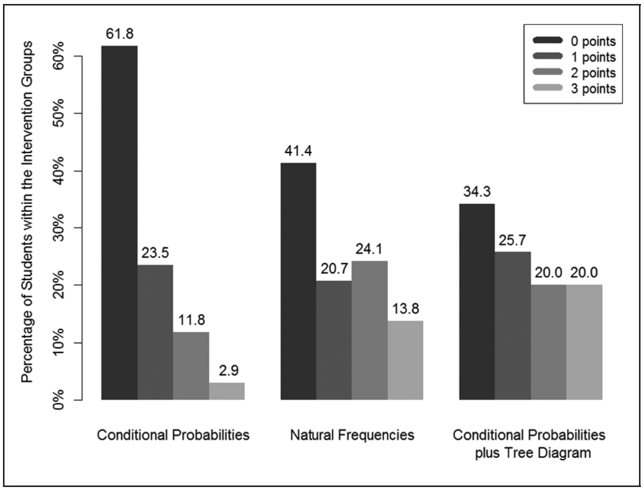

Figure 2 Distribution of student scores within groups.

need to be repeated in order to strengthen these skills.[21] However, this is difficult in the already crowded curriculum in medical studies. Translation of conditional probability formats into natural frequencies has been shown in this and other studies to improve participants' test results.[9] However, according to Hoffrage and Gigerenzer,[14] people do not spontaneously translate probabilities to natural frequencies. Furthermore, in the process of this conversion, miscalculations can occur. In light of the results of our study, we suggest that in future students as well as physicians use the tree diagram as an aid for numerical processing when dealing with diagnostic tests expressed in conditional probabilities.

REFERENCES

1. Sheridan SL, Pignone M. Numeracy and the medical student's ability to interpret data. Eff Clin Pract. 2002;5(1):35–40.

2. Windish DM, Huot SJ, Green ML. Medicine residents' understanding of the biostatistics and results in the medical literature. JAMA. 2007;298(9):1010–22.

3. Gramling R, Irvin JE, Nash J, Sciamanna C, Culpepper L. Numeracy and medicine: key family physician attitudes about communicating probability with patients. J Am Board Fam Pract. 2004;17(6):473.

4. Wulff HR, Andersen B, Brandenhoff P, Guttler F. What do doctors know about statistics? Stat Med. 1987;6(1):3–10.

5. Berwick DM, Fineberg HV, Weinstein MC. When doctors meet numbers. Am J Med. 1981;71(6):991–8.

6. Estrada C, Barnes V, Collins C, Byrd JC. Health literacy and numeracy. JAMA. 1999;282(6):527.

7. Kurzenhauser S, Hoffrage U. Teaching Bayesian reasoning: an evaluation of a classroom tutorial for medical students. Med Teach. 2002;24(5):516–21.

8. Noguchi Y, Matsui K, Imura H, Kiyota M, Fukui T. Quantitative evaluation of the diagnostic thinking process in medical students. J Gen Intern Med. 2002;17(11):839–44.

9. Hoffrage U, Kurzenhauser S, Gigerenzer G. [How can one improve the understanding and communication of the importance of medical test results?]. Z Arztl Fortbild Qualitatssich. 2000;94(9):713–9.

10. Eddy DM. Probabilistic Reasoning in Clinical Medicine: Problems and Opportunities. Cambridge (UK): Cambridge University Press; 1982.

11. Rao G, Kanter SL. Physician numeracy as the basis for an evidence-based medicine curriculum. Acad Med. 2010;85(11):1794–9.

12. Freeman JV, Collier S, Staniforth D, Smith KJ. Innovations in curriculum design: a multi-disciplinary approach to teaching statistics to undergraduate medical students. BMC Med Educ. 2008;8:28.

13. Noguchi Y, Matsui K, Imura H, Kiyota M, Fukui T. A traditionally administered short course failed to improve medical students' diagnostic performance: a quantitative evaluation of diagnostic thinking. J Gen Intern Med. 2004;19(5 Pt 1):427–32.

14. Hoffrage U, Gigerenzer G. Using natural frequencies to improve diagnostic inferences. Acad Med. 1998;73(5):538–40.

15. Gigerenzer G, Hoffrage U. How to improve Bayesian reasoning without instruction: frequency formats. Psychol Rev. 1995;102(4):684–704.

16. van der Windt DA, Simons E, Riphagen II, et al. Physical examination for lumbar radiculopathy due to disc herniation in patients with low-back pain. Cochrane Database Syst Rev. 2010;(2): CD007431.

17. Cohen J. A power primer. Psychol Bull. 1992;112(1):155–9.

18. Galesic M, Garcia-Retamero R. Graph literacy: a cross-cultural comparison. Med Decis Making. 2011;31(3):444–57.

19. Garcia-Retamero R, Galesic M. Who profits from visual aids: overcoming challenges in people's understanding of risks [corrected]. Soc Sci Med. 2010;70(7):1019–25.

20. Edwards A, Elwyn G, Mulley A. Explaining risks: turning numerical data into meaningful pictures. BMJ. 2002;324(7341): 827–30.

21. Astin J, Jenkins T, Moore L. Medical students' perspective on the teaching of medical statistics in the undergraduate medical curriculum. Stat Med. 2002;21(7):1003–6; discussion 1007.